

Hand-in 4: XML Processing

The hand-in should be sent, by the same groups as in the first hand-ins, to Andrea Campagna, acam@itu.dk at the latest

Monday November 10, 23.59.

Parts of the hand-in can be written by hand and scanned, but it has to be handed in electronically.

*The hand-in will be corrected and you will get feedback from Andrea. The feedback is an **important** part of your learning process.*

The fourth part of the project is loosely connected to the previous three parts. However, together with the fourth part you should hand in any revision of the previous hand-ins that you may have, as agreed with Andrea.

You will be working with the file `actors.xml`, downloadable from the course home page. It contains information on actors and movies from IMDB similar to the information that you worked with previously.

XML Schema

Your first task is to make an XML schema, *ActorML*, that describes the data format in `actors.xml`. Your schema should be as specific as possible, i.e., not allow significant deviations from the format and conventions used. Make reasonable assumptions. Run a validator to ensure that `actors.xml` is contained in the language described by the schema. In the same way, you should make an XML Schema for a markup language *MovieML* with information about *movies*, including at least title, year, and actors starring in the movie.

XML Queries

Write the following queries in XQuery, and run them on `actors.xml`:

1. Which movies have John Travolta and Uma Thurman starred in together?
2. What is the name of all actors in Pulp Fiction? Your query should list the actors sorted by first name.
3. What are the titles and years of all movies from the 1990s where Tom Hanks starred?

XML Transformation

You should write an XSLT stylesheet that transforms XML documents in ActorML format into XML documents in MovieML format. You should test your stylesheet on `actors.xml`. In particular, you should run the output of your stylesheet on a validator to test conformance with your XML Schema.

To be handed in

- Name of all group members, including e-mail addresses
- Your XML schemas, including a description of any nontrivial choices made
- Your XSLT stylesheet
- Your XQuery queries
- If you are not satisfied with some part of the hand-in, e.g. if you have discussed alternative solutions and don't know which is the best choice, you should add some comments on it.

In addition, you should be prepared to demonstrate your program in connection with the exercises.

Course goals covered by this hand-in

After the course the students should be able to:

- write simple XML Schemas and simple XQuery.
- explain the meaning of a DTD, and the effect of simple XSLT transformations.